

HDMI Modeling Design Doc

gdn-cpa-core@,
(Original: July 2015 OPC)
(Updated Aug 2015 for MPC, ROAS)
(Updated Oct 2015 for auto/manual split)
(Updated May 2018 for low-data)
(Updated Sep 2019 for store-visit conversions)

Ref: Modeling Productionisation Guide

go/gdn-co-hdmi-modeling-design-doc

General Information

1. What does the model predict?

- a. [REDACTED]
 - i. [REDACTED]
 - 1. [REDACTED]
 - ii. [REDACTED]
- b. [REDACTED]
 - i. [REDACTED]
 - ii. [REDACTED]
- c. Updated model prediction description as of 2018-05-10
 - i. [REDACTED]
 - 1. [REDACTED]
 - 2. [REDACTED]
 - 3. [REDACTED]
 - 4. [REDACTED]
 - 5. [REDACTED]

2. How and where is the prediction used?

- a. [REDACTED]
[REDACTED]

3. Contact information (team/TL email, wiki page).

- a. TLM: adarvish@
- b. team: gdn-cpa-core@
- c. web: go/project-hdmi

Commented [1]: how are these [REDACTED] chosen?

Commented [2]: There's a [REDACTED] section below.

Source of Training Data

1. How is the training data obtained?

- a. [REDACTED]
 - b. If it is based on [REDACTED]
 - i. N/A
 - c. If it reads from [REDACTED], have you checked for the logs being ready before consuming them?
 - i. N/A
2. Does it correctly handle conversion data (see Conversion Handling)?
- a. [REDACTED]
 - b. [REDACTED]
 - c. [REDACTED]
3. Does it correctly handle DBM data (see DBM Data Separation)?
- a. N/A, since the data source if [REDACTED]
4. How are bad data handled (e.g., when certain days of data are deemed unusable)?
- a. [REDACTED]
5. Are the training data stored?
- a. If so, does it satisfy wipeout compliance? [REDACTED]
6. How frequently does the training data get refreshed?
- a. [REDACTED]
7. What happens when training data is stale?
- a. [REDACTED]

Commented [3]: update?

Commented [4]: Not changed in this launch but in need of update. Started one below.

Commented [5]: +kehuang@google.com
+sunder@google.com +nouri@google.com

Updated the model design doc for store visit conversions.

Training Algorithm

1. How is the model trained?
 - a. If it is based on existing infrastructure [REDACTED]
 - i. N/A
 - b. If it is a custom learning infrastructure,
 - i. [REDACTED]
 - ii. What is the canary plan when the learning infrastructure is changed?

1. [REDACTED]
2. How frequently is the model updated?
 - a. If it is custom and updated on a regular basis, <link to [REDACTED]
 - i. [REDACTED]
 - b. If it is on demand, how do we decide a new model is needed?
 - i. N/A
3. What is the metric for evaluating model quality?
 - a. [REDACTED]
4. Are the models versioned?
 - a. [REDACTED]

Serving

1. How is the model served?
 - a. [REDACTED]
 - b. Update 2018-05-10. [REDACTED]
2. How does the model get updated in serving?
 - a. Is there a canary process for the new model?
 - i. [REDACTED]
 - ii. [REDACTED]
3. In case of production issues,
 - a. How do we rollback to an older version of the model?
 - i. [REDACTED]
 - ii. Updated 2018-05-10. [REDACTED]
 - b. Can we disable the model?
 - i. [REDACTED]
4. Are there [REDACTED] involved in serving (such as thresholds)?
 - a. How are they selected?
 - i. Yes, fallback constants and thresholds:
 1. [REDACTED]
 2. [REDACTED]

Commented [6]: what is this comparison?

Commented [7]: Not changed in this launch but in need of update.

ii. [REDACTED]

Commented [8]: are this the old ones? or are they updated?

Commented [9]: See "Update" halfway through selected text.

b. How do they get updated?

i. [REDACTED]

ii. Update 2018-05-10.

1. [REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]

2. [REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]

Commented [10]: how would you know that it's time to tune them? are there alerts?

Commented [11]: Discussed with various stakeholders in the launch doc. We plan on doing a post-experiment analysis on this to make sure that nothing's getting caught in the rampup state.

Commented [12]:

Commented [13]: [REDACTED]

Commented [14]: wait, are you saying that gdn bidding has no call at all that can be paged in emergencies?

Commented [15]: Outdated. I updated parts marked updated. gdn-cpa-oncall@ gets pages.

Commented [16]: are these email pages or actual phone call pages?

Commented [17]: Calls.

Monitoring

1. Which group is responsible for monitoring?

a. `gdn-cpa-alerts@google.com`

b. `gdn-cpa-oncall@` gets pages.

2. (This is not an exhaustive list) Are there alerts for

a. Training data looks bad and/or incomplete

i. [REDACTED]

- b. Trained models have poor quality

i. [REDACTED]
[REDACTED]
[REDACTED]

- c. Trained models are too different from the last iteration
 - i. [REDACTED]
- d. Models are stale
 - i. [REDACTED]
- e. Jobs failures (job down, [REDACTED] overrun, [REDACTED] failures)
 - i. Yes, [REDACTED] will email
 - ii. [REDACTED]
 - iii. [REDACTED]
- f. Server side monitoring (see productionisation guide for suggestion)
 - i. [REDACTED]
 - ii. We plan on setting up the following alerts, once we are able to verify reasonable thresholds via [REDACTED]:
 - 1. [REDACTED]
 - 2. [REDACTED]
 - 3. [REDACTED]
- 3. Is there a playbook describing how to handle the different types of alert?
 - a. go/hdmi flume alert playbook

Production Readiness

- 1. How are PCRs handled?
 - a. Is training multi-homed?
 - i. [REDACTED]
 - ii. go/hdmi-pipeline-dc-migration (by jtakagi@)
- 2. Does it require any production resources?
 - a. If so, please review usage on a regular basis and make plans for resources well in advance
 - i. [REDACTED]
 - ii. [REDACTED]
- 3. If it uses batch resources, is there monitoring of usage against batch quota?
 - a. [REDACTED]

4. In case of resource crunch,

a. Are there ways to temporarily resolve the issue [REDACTED]

i. [REDACTED]

b. Are there ways to reduce load by turning off jobs of lower priorities (for example,